UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER BBN Report No. 3209	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMMAND AND CONTROL RELATED COMPUTER TECHNOLOGY		5. TYPE OF REPORT & PERIOD COVERED Quarterly Progress Report 1 Sept. 75 - 31 Nov. 75
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) J. D. Burchfiel R.S. Nickerson R. Viswanathan A.W.F. Huggins		8. CONTRACT OR GRANT NUMBER(s) MDA903-75-C-0180
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Bolt Beranek and Newman Inc. 50 Moulton Street, Cambridge, Mass. 02138		12. REPORT DATE December 1975
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce for sale to the general public.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This research was supported by the Defense Advanced Research Projects Agency under ARPA Order No. 2935.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

packet radio, computer communications, TENEX TCP, PDP-11 TCP, station gateway; ELF, BCPL, cross-network debugging, speech compression, vocoder, linear prediction, covariance lattice, spectral distance, speech-quality evaluation, multidimensional scaling, MDPREF, INDSCAL

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This document describes progress on (1) the development of a packet radio network, (2) speech compression, and (3) speech-quality evaluation. Activities reported under (1) include work on PDP-11 and TENEX TCP development, station gateway and ELF development, LAD-4 forwarder and digital unit checkout; under (2) spectral distance measures of vocoder performance and investigation of covariance lattice methods of linear prediction; and

(continued over)

20. (continued)

under (3) correlation of multidimensional scaling results with rating and ranking data, and experimentation with phoneme-specific quality tests.

TABLE OF CONTENTS

I. INTRODUCTION

Packet Radio development during this quarter has produced significant interaction between areas which had been largely separate until now. Interaction of effort areas within BBN, such as TCP development and gateway development, has necessitated resolution of the interfaces between Packet Radio station modules. As these interfaces are defined in detail, implementation progresses and coding and joint debugging follow.

The interaction of effort areas extends to the diverse tasks assigned to Packet Radio contractors. During this quarter, Collins Radio delivered the initial version of software whose level of operation directly impacts station algorithms, structure and operation. Detailed analysis of their work, and its relation to the design efforts already in progress or completed at BBN, has been a major challenge. Through frequent contact with Collins and other contractors, the station design is moving toward the specificity of an initial implementation. In sections below various aspects of this process are discussed. The TCP and gateway, for example, are in advanced stages of debugging. The control and connection processes, on the other hand, are well understood but hinging on resolution of protocol and operational details.

The meeting of previously distinct effort areas has proven a rich ground for auxiliary development. The cross-network debugger has been particularly enriched as its use intensified. The station operating system has reached operational stability. The

Transmission Control Program development in the PDP-11, which will provide the PR network user access to other networks, is proceeding synergistically with TCP development in the PDP-10, where the other end of the user's connection will reside.

II. MEETINGS AND PUBLICATIONS

During this quarter BBN issued PRTN 156, "Gateway Design for Computer Network Interconnection." This document describes the functional role played by a gateway between two computer networks. General issues inherent in the interconnection of two potentially different networks are discussed. Attention then focuses on implementation questions crucial to installation of a gateway on the Packet Radio network. Concepts such as the Gateway Virtual Network, and the Gateway Coordination Center are introduced. The document provides a fundamental description of gateway issues from which finer details may grow.

In BBN work during this quarter progress was made both in gateway design and Transmission Control Program (TCP) implementation. The two areas of development met, with the result that a document describing the interface between a TCP and a gateway was prepared. This was issued informally and circulated to interested parties.

This quarter also saw the distribution of critiques reviewing Channel Access Protocol documents, as described below in section VIII.

Documents describing the source, history, status, and use of support software, station operating system, and test programs were delivered to SRI staff visiting BBN during the last week of October. During this visit several areas of Packet Radio work were discussed,

including protocols, station design, and development priorities. Avenues for cooperative work between SRI and BBN were also explored. Earlier in October other SRI personnel visited BBN for work on the ELF operating system. As a result of this meeting and previous cooperative efforts, final bugs were worked out of the system and a final consensus on modifications was reached. ELF work is treated in greater detail in section VI below.

III. PDP-11 TCP DEVELOPMENT

During this quarter, the PDP-11 TCP was brought to the point of having most of the bugs introduced by the conversion procedure from the TENEX version eliminated and final checkout initiated. The remaining tasks are to incorporate the bug-fixes made to the TENEX version, integrate the TCP with Gateway, checkout under a variety of conditions, make performance measurements, and conduct a variety of experiments.

IV. TENEX TCP DEVELOPMENT

The basic Transmission Control Program [Kahn _ Cerf, International Network Working Group Memorandum No. 39] has been implemented and debugged. Extensive metering and internal statistics taking have been included. These have proved invaluable in both packet throughput measurements and in finding bottlenecks in the TCP code.

The TENEX TCP has been tested by communicating with the TCP at Stanford University Digital Systems Laboratory. During these sessions several deficiencies in the protocol were discovered. In particular, the protocol for closing a connection (the FIN control function) was inadequate. This issue was resolved in meetings between Cerf of SU-DSL and Tomlinson of BBN. The TENEX TCP now uses the revised version of the protocol.

One of the main features of the TCP is its ability to periodically resynchronize connections in order to guarantee that the sequence numbers used in packets will never conflict with any which might be in use in packets or their duplicates which could be reverberating in the network. This has been thoroughly tested and demonstrated to work if each end of the connection independently resynchronizes, and if both ends simultaneously resynchronize.

Several auxiliary programs have been written which use the Internet Protocol. ECHO is a simple program which echos messages sent to it. TTLSRV is a simple TCP Telnet server which allows

logging into BBNA from a remote site using the Internet Protocol. TTLUSR is the user companion to TTLSRV -- it is allows terminal communications with a TCP at a remote site which is running TTLSRV.

TCPTST is a test program which opens a TCP connection to itself. One process in the program serves to send Internet letters over the connection which are received by another process. In manual mode these letters are simply lines of text typed on the terminal (very similar to an echo program). In automatic mode the letters are produced by the program. The receiving process checks for errors in sequencing and data in automatic mode.

GATEWAY is a very simple gateway program which runs on BBNB. Operationally, GATEWAY forwards internet packets received on ARPANET link 155 onto link 158 and vice versa. This permits experimenters to view link 155 as "ARPANET A" and link 158 as "ARPANET B" (Internet numbers 12 and 13). Two TCPs may be operated, one in each logical network. Messages destined for a host in the other network are sent to the GATEWAY program for forwarding.

Current activity in the TCP is directed towards improving the packet throughput which is currently only approximately 23 packets per second. The largest increase in throughput was achieved by hand coding only about a dozen routines and resulted in a 146 per cent increase. It is expected that improvements in the basic TCP algorithms will result in a significant increase in the future.

V. STATION GATEWAY DEVELOPMENT

The station gateway program has been coded and partially debugged. Currently, the only function provided by the gateway is the re-addressing of packets. Each packet sent through the gateway contains an internet header which is used by the gateway to form a local header for the destination network. As the source and destination networks may be the same, we were able to debug the portion of the gateway dealing with the ARPANET. The program has been used as a gateway on the ARPANET handling traffic between the TCP on the PDP-10 and the TCP on the PDP-11. This required specification and coding of the interface between the TCP on the PDP-11 and the gateway. The sections of the gateway dealing with traffic between the Packet Radio Net and the Arpanet have been coded but cannot be debugged until the connection process which interfaces the gateway to the network is written.

VI. ELF DEVELOPMENT AND PERFORMANCE

We have worked with Dave Retz to produce a version of the ELF kernel which includes all of the modifications made by BBN since the virtual system was released last May. This version of the kernel has been in use at BBN and at SRI since October with no modifications and is due to be released to the ELF community by the end of December. We have also studied the performance of the ELF system and have made several suggestions on how to improve its efficiency. The performance issues are currently under consideration by Dave Retz.

VII. BCPL RUNTIME SUPPORT

The BCPL runtime library has been expanded to provide routines for handling semaphores and timers, and to provide additional functions in the I/O routines. The library has been re-coded to improve the efficiency of teletype I/O by reducing the number of processes and the number of copy operations on buffers needed to do I/O. The new implementation has been debugged and is being used by the TCP and Gateway programs.

VIII. CONNECTION PROCESS

The major developments in the design of the Packet Radio Station's Connection Process during this quarter are conceptual in nature. The underlying assumptions concerning basic operation of the Packet Radio Network made by each contractor are now interacting with the implementation efforts of the other contractors. BBN is taking an active role in identifying and resolving these issues.

Considerable effort was expended this quarter on review, analysis and suggestions for revision of protocol released by Collins Radio as PRTN #144-R1, "Packet Radio Channel Access Protocol Program." BBN participated in the review both of this document and of comments from Stanford Research Institute. The upshot of this effort was a more thorough understanding of protocol issues, and consequently a more exact basis upon which to implement the connection process in the station.

IX. DESIGN OF CONTROL PROCESS

Work began recently on design of the station control process, which is responsible for labeling (determining how packets are to be routed through) the network. It must not only assign initial labels, but also relabel, if necessary, as network connectivity changes. The design of this program of course depends strongly on details of network behavior. Thus we have devoted much of our effort so far to understanding and trying to improve the protocols which govern the generation and forwarding of packets by packet radio units. We studied and wrote a critique of PRTN 144, Collins' documentation of the PR program for initialization and packet routing, and responded to SRI's critique of the same. The issues raised in this critique, as well as other problems, were discussed at our December 4 meeting with Collins and SRI. Many questions were resolved, but some require further thought; we will be working closely with Collins and the other contractors to decide them.

In designing the station control process, we must understand such questions as:

- What is the routing protocol; what are labels?
- What information pertinent to labeling is available to the station and how is it obtained?
- How does the station evaluate this information; what decisions does it make?
- What inconsistencies can arise; how can they be minimized and dealt with?

Our approach is to first design a program that will work in an idealized network, making simple decisions based on easy-to-obtain

information. The design will then be expanded along each of these fronts: to take complexities of real network behavior into account; to make more intelligent decisions; and to utilize more sources of information. This approach should produce a "working" program in a short time, which will then evolve into the final program.

X. LAD-4 DEMONSTRATION FORWARDER

The simple forwarder to be used for LAD-4 was demonstrated to Mike Placko of SRI in October. Currently, the forwarder is run in a loop mode in which the forwarder generates packets which are looped back to the PDP-11 by the PRU through the DMA interface. This has enabled us to debug the forwarder, and it is now ready for further check out pending Collins release of the CAP software needed by the PRUs to communicate with the station forwarder. The forwarder and documentation covering both the user interface and internal operation of the program have been delivered to SRI.

XI. CROSS-NET DEBUGGER

Substantial improvements were made to the PDP-11/TENEX cross-net debugger during this quarter. These improvements enhance the ability to apply the power of TENEX to the task of debugging programs running on a PDP-11. The initial development of this facility has been described in earlier reports. The improvements described in this report fall into four areas: Debugging multiple processes, symbolic debugging, diversion of PDP-11 console TTY output to TENEX, and user conveniences.

The goal in developing the facility for debugging multiple processes was to permit a user or users to be actively debugging more than one process simultaneously. Furthermore, the restriction of the debugger to a single host needed to be removed. An example where this facility is needed is in the case where a new program is being exercised by a test program. One debugger is used on the program being debugged, and the other is used to run the exerciser. First, the debugger was made to accept messages from all hosts on the debugger link and then rejecting those which are not contained in a list of hosts permitted to debug that particular PDP-11. This allows debugging from more than one host; however, it is not yet possible to distinguish between different uses from a single host. Another field of the header (the 4-bit extension of the message ID) is used in addition to the host field to select a logical connection from which the process being debugged and associated information can be found. A corresponding change in the TENEX X-NET to select an

network to the debugger on TENEX which prints the text on the TENEX terminal and sends a reply indicating its completion to the debugger in ELF. The debugger in ELF then signals the completion of the output just as if the original SIO had been executed. This facility will be extended to the case of console input as well.

Several conveniences have been added to X-NET. Generally, these fall into the class of providing a short command for performing a frequently occurring sequence of actions. An example is that of connecting to a PDP-11, creating a process, and loading the memory of that process with the image stored in the X-NET. Another example is the ability to execute an arbitrary instruction in the PDP-11 and have control returned to the TENEX debugger upon its completion. This avoids the necessity of depositing the instruction in the PDP-11, planting a breakpoint following it, remembering the current contents of the program counter register, starting the instruction, and then restoring the program counter register to its original value.

XII. DIGITAL UNIT ATTACHMENT AND TEST

During this quarter, the first Packet Radio Digital Unit has completed the checkout necessary to assure its correct operation. The random access memory board identified as failed during the preceding quarter was returned to Collins Radio, repaired, and reinstalled in PRDU #1. The diagnostics detected what at first seemed to be an error, but consultation with Collins Radio and SRI located a discrepancy between the mode of operation of the diagnostic in question and the PRDU design. After modification of the diagnostic, PRDU #1 ran all testing and exercising programs without error.

PRDU #2 arrived at BBN during this quarter. It was unpacked and placed in the BBN Packet Radio area, where diagnostics were run. Low level diagnostics ran successfully. An obscure problem arose when sending packets to PRDU #2 from the PDP-11, however. It appears that packets longer than some critical length cause the receive DMA interface to malfunction in some way which halts the PRDU. The circumstances of the failure are peculiar, such as that the new Link Test Support program, and not the old, must be running in the PRDU. We are in contact with Collins Radio on this problem, and expect that problem identification and resolution will follow.

TABLE OF CONTENTS

I. INTRODUCTION

During the last quarter we made progress in the areas of objective speech quality evaluation and real-time implementation of the LPC algorithm.

Based on published subjective results dealing with perceptual difference limens for formant frequencies of vowels, we developed a procedure for checking the validity, or what we call perceptual consistency, of a given objective spectral distance measure. As an important result, we derived a general necessary condition for perceptual consistency for a class of spectral distance measures. We experimentally investigated a number of distance measures developed in the previous quarter to check for their perceptual consistency. Although none of the measures developed to date was found to be perceptually consistent for all (frequency) locations of formants, our experiments have suggested the use of new weighting functions in forming spectral distance measures towards achieving perceptual consistency.

We detected and corrected a number of significant hardware problems in our SPS-41 machine. Also, the real-time LPC algorithm was demonstrated to run on our machine in back-to-back mode with an acceptable mean time between failures.

Also in the last quarter we have begun work in developing a new class of stable linear prediction methods, which we have called lattice covariance methods.

II. SPECTRAL DISTANCE MEASURES AND PERCEPTUAL CONSISTENCY

In our formulation of the problem of objective evaluation of LPC vocoder speech quality, we compare the unquantized and the quantized (and interpolated) sets of filter parameters computed over a number of data frames [1]. For any given data frame, the problem is to determine the deviation or distance between the two sets of filter parameters, or equivalently between the corresponding linear predictor spectra. In the previous quarter, we developed a number of spectral distance measures for this purpose [1]. The question, then, is how to evaluate the validity of a given spectral distance measure. Inasmuch as the vocoded speech is to be perceived by human listeners, it is appropriate to require of a distance measure to be at least consistent with the known properties of human speech perception.

As a first step, we have used Flanagan's perceptual results on difference limens for vowel formant frequencies [2] as one basis for checking the perceptual consistency of distance measures. Briefly summarizing Flanagan's work, when two formants are in close proximity, human perception exhibits an asymmetrical pattern in that moving one of the two formants closer to the other by a given amount produces a larger perceived quality difference than moving that formant away from the other by the same amount. On the other hand, the same formant shifts produce a symmetrical pattern when the two formants are well separated.

Below we first present the types of spectral distance measures we have investigated. We then state and prove a necessary condition for perceptual consistency of a class of these distance measures. Finally, we give the results of our experimental study of a number of previously developed distance measures.

A. Spectral Distance Measures

A large class of spectral distance measures between the reference and the test spectra $P_r(\omega)$ and $P_t(\omega)$ is defined by the L_k norm:

$$d_k(P_r, P_t, W) = \left[\frac{\int_{-\pi}^{\pi} W(P_r(\omega),\ P_t(\omega),\ \omega)\ |e(\omega)|^k\ d\omega}{\int_{-\pi}^{\pi} W(P_r(\omega),\ P_t(\omega),\ \omega)\ d\omega} \right]^{\frac{1}{k}} , \qquad (1)$$

where W is a positive weighting function that depends, in general, on $P_r(\omega)$, $P_t(\omega)$ and frequency ω ; $e(\omega)$ is the error between normalized versions of $P_r(\omega)$ and $P_t(\omega)$. In our investigations, we have mainly used arithmetic mean (AM) normalization or geometric mean (GM) normalization [1]. Examples of the error function are:

$$e(\omega) = P_r(\omega) - P_t(\omega) , \qquad (2)$$

$$e(\omega) = \log P_r(\omega) - \log P_t(\omega) = \log\left[P_r(\omega)/P_t(\omega)\right] , \qquad (3)$$

$$e(\omega) = P_r(\omega)/P_t(\omega) , \qquad (4)$$

where the spectra are assumed to have been normalized.

B. A Necessary Condition for Perceptual Consistency

Fig. 1 shows two plots of spectral deviation or distance versus frequency shift of the second formant causing that spectral deviation. (Frequencies of the other three fixed formants and the nominal value of the second formant frequency are given in the figure. Bandwidths of the formants were fixed at 130 Hz, 150 Hz, 185 Hz, and 200 Hz, respectively for the first, second, third and fourth formants [2].) Fig. 1(a) corresponds to the error definition (4) while Fig. 1(b) corresponds to the error definition (3). These are two of the most popular error definitions being used for computing spectral distance measures in many speech processing applications. Both plots in Fig. 1 were obtained using GM normalization, k=1 and no weighting in (1). (We have plotted log d_1 for plot (a) so that ordinates of both plots are in decibels.) The almost symmetrical plots in Fig. 1 do not conform with properties given by Flanagan (see Fig. 4(c) in [2]).

Notice that the two distance measures that produced the plots in Fig. 1 depend only on the ratio of the spectra P_r and P_t (in view of (3) and (4)). Below we prove that with GM normalization, any distance measure which is a function of only the ratio of the spectra is necessarily perceptually inconsistent. First, we give our working definition of perceptual consistency, based on Flanagan's results [2].

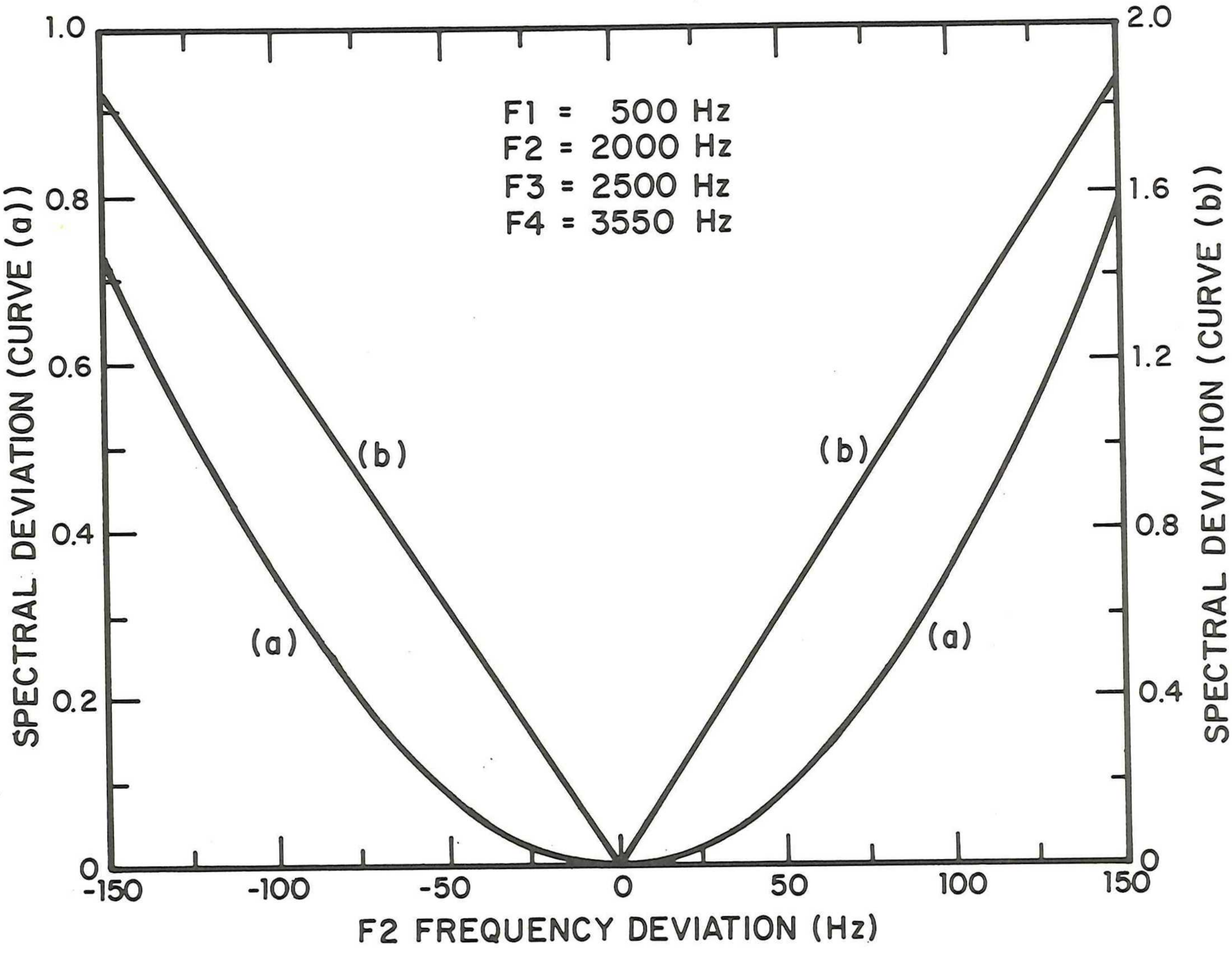


Fig. 1. Plots of spectral deviation versus shift in second formant frequency for two spectral error definitions.

<u>Working Definition of Perceptual Consistency</u>: Let X and Y be two vowel spectra, such that Y is identical to X except that one of the formant frequencies F is shifted by a variable amount ΔF. A given spectral distance measure $d(X,Y)$ between X and Y is said to be perceptually consistent if

(a) when F is close to another formant F', $d(X,Y)$ exhibits asymmetry such that it is greater when F is moved ΔF towards F', than when F is moved ΔF away from F';

(b) such asymmetry decreases as F and F' are further apart.

Now, consider a class $\mathcal{D}_g$ of spectral distance measures defined by (1) where the error $e(\omega)$ is computed after GM normalization of the spectra. For this class of distance measures, a necessary condition for perceptual consistency is provided below in the form of a theorem.

<u>Theorem</u>: A necessary condition for any spectral distance measure $d(P_1,P_2)$ in the class $\mathcal{D}_g$ to be perceptually consistent (as defined above) is that it not be a function of only the ratio of the two spectra P_1 and P_2.

<u>Proof</u>: Assume that a distance measure in $\mathcal{D}_g$ violates the necessary condition. We show that this distance measure is not perceptually consistent. Let P_2 be obtained from P_1 by shifting only one of its formant frequencies while keeping all other parameters intact. Let the denominator of the linear predictor power spectrum ($S(\omega)$ in Eq. (1) of [1]) be factored into $R(\omega)$ and $S'(\omega)$, where $R(\omega)$ is the

contribution to the spectrum from the formant under consideration and $S'(\omega)$ represents the contributions from all other poles of the linear predictor. Thus, $P_1(\omega) = 1/(R_1(\omega).S_1'(\omega))$ and $P_2(\omega) = 1/(R_2(\omega).S_1'(\omega))$, where $R_2(\omega)$ is the perturbed version of $R_1(\omega)$. This gives the result that the ratio of P_1 and P_2 depends only on the formant under consideration. Specifically, the ratio does not depend on whether or not this formant is in close proximity to another formant. This clearly establishes that the measure is not perceptually consistent according to our working definition.

C. Experimental Results

We tested all the distance measures that we developed in the previous quarter [1] for their perceptual consistency. We employed a number of reasonable frequency weighting functions based on spectral intensity, articulation index, and frequency derivative of spectrum.

Several specific results were obtained in this study. First, even with AM normalization, when the necessary condition stated in the last section was violated, perceptual consistency was not obtained. Second, inasmuch as one is looking for sensitivity to formant interaction, spectral distance is best defined using GM normalization and with spectral error defined as the difference in the (linear) spectral domain as given in (2).

All the spectral distance measures that we investigated, even with the use of the above-mentioned weighting functions, had one

common problem in that for the case when the first formant frequency was shifted about the nominal value of 300 Hz, a given amount of left shift always produced a larger spectral deviation than a right shift of the same amount, which is just the opposite of what Flanagan reported (see Fig. 3(a) in [2]). (We found, however, that some of these measures and weighting functions produced the right types of asymmetry in other test conditions considered by Flanagan.) To attempt to overcome this problem, we are currently investigating the use of weighting functions based on perceived loudness of speech [3].

III. REAL-TIME IMPLEMENTATION

For most of the last quarter, our SPS-41 machine was at SPS Inc. in Waltham for purposes of hardware debugging. During that time, the machine was tested extensively. Several redesigned boards were tested. In addition, several faulty chips were found and replaced. With the modifications installed, several versions of the back-to-back LPC software were run for a considerable length of time. Several runs of over eight hours were observed, and most runs were terminated after several hours without failing. SPS Inc., then, built two complete sets of boards, which when substituted into our machine, resulted in the same level of performance. These new boards, which will be available to the ARPA sites, should simplify greatly the problems of debugging machines in the field.

Our SPS-41 machine was returned and installed at BBN. It is currently being tested to see if the good performance can be repeated here at BBN. When it is found to perform satisfactorily, the exact versions of the programs used will be made available to the other ARPA sites for whatever debugging they may choose to do.

IV. COVARIANCE LATTICE METHODS OF LINEAR PREDICTION

The autocorrelation method of linear prediction guarantees the stability of the all-pole filter, but has the disadvantage that windowing of the signal causes some unwanted distortion in the spectrum. In practice, even the stability is not always guaranteed with finite wordlength (FWL) computations. On the other hand, the covariance method does not guarantee the stability of the filter, even with floating point computation, but has the advantage that there is no windowing of the signal. One solution to these problems was given by Itakura in his lattice formulation. In this method, filter stability is guaranteed, with no windowing, and with FWL computations. Unfortunately, this is accomplished with about a four-fold increase in computation over the other two methods.

We have begun investigation to develop lattice methods which have all the properties of a regular lattice but where the number of computations is comparable to the autocorrelation and covariance methods. In these methods the"forward" and "backward" residuals will not be computed. The reflection coefficients will be computed directly from the covariance of the input signal. Thus, the new methods will be a hybrid between the covariance method and traditional lattice methods. This is why we have called these methods covariance lattice methods.

REFERENCES

1. BBN Quarterly Progress Report on Command and Control Related Technology, Part II, Report No. 3122, Sept. 1975.

2. J.L. Flanagan, "A Difference Limen for Vowel Formant Frequency," J. Acoust. Soc. Amer., Vol. 27, No. 3, pp. 613-617, May 1955.

3. S.S. Stevens, "Perceived Level of Noise by Mark VII and Decibels (E)," J. Acoust. Soc. Amer. , Vol. 51, No. 2, pp. 575-600, Feb. 1972.

TABLE OF CONTENTS

I. INTRODUCTION

Our work on evaluation of vocoded speech quality has proceeded along two main lines during the last quarter. We have carried out more detailed analyses of the data collected with the rating procedure described in QPR #3, and correlated these results with parallel analyses of the data collected with the rank ordering task, described in QPR #2. We have completed a pilot experiment of the phoneme-specific test procedure, and have used the results to improve the procedure, and increase the amount of information obtained from the test. We have continued work on our review paper on quality-assessment techniques, and a paper on some of our subjective quality results was presented at the 90th meeting of the Acoustical Society of America in November.

II. FURTHER ANALYSIS OF THE RATING AND RANKING DATA

A. Three Dimensional MDPREF Analysis.

The data from the rating task (see QPR #3 for details) consist of a 36 x 14 matrix. Each cell of the matrix contains a figure that represents the combined ratings given by the four subjects to a particular sentence spoken by a particular talker, processed through a particular vocoder system. The 36 rows of the matrix correspond to the 36 speaker x sentence combinations, and the 14 columns to the 14 vocoder systems used in the tests. The multi-dimensional scaling program MDPREF attempts to model the input data by treating the vocoder systems as points in an N-dimensional space, and the

speaker-sentence combinations as vectors through the space. To obtain the modelled ratings of the vocoder systems, for a particular speaker-sentence combination, the points representing the systems are projected onto the vector representing the speaker-sentence combination, and the ratings are given by the positions of the projected points on the vector. (See Fig. 1) The analysis finds the positions of the points and vectors in the N-dimensional space that gives the best least-squares fit between the empirical and the modelled data.

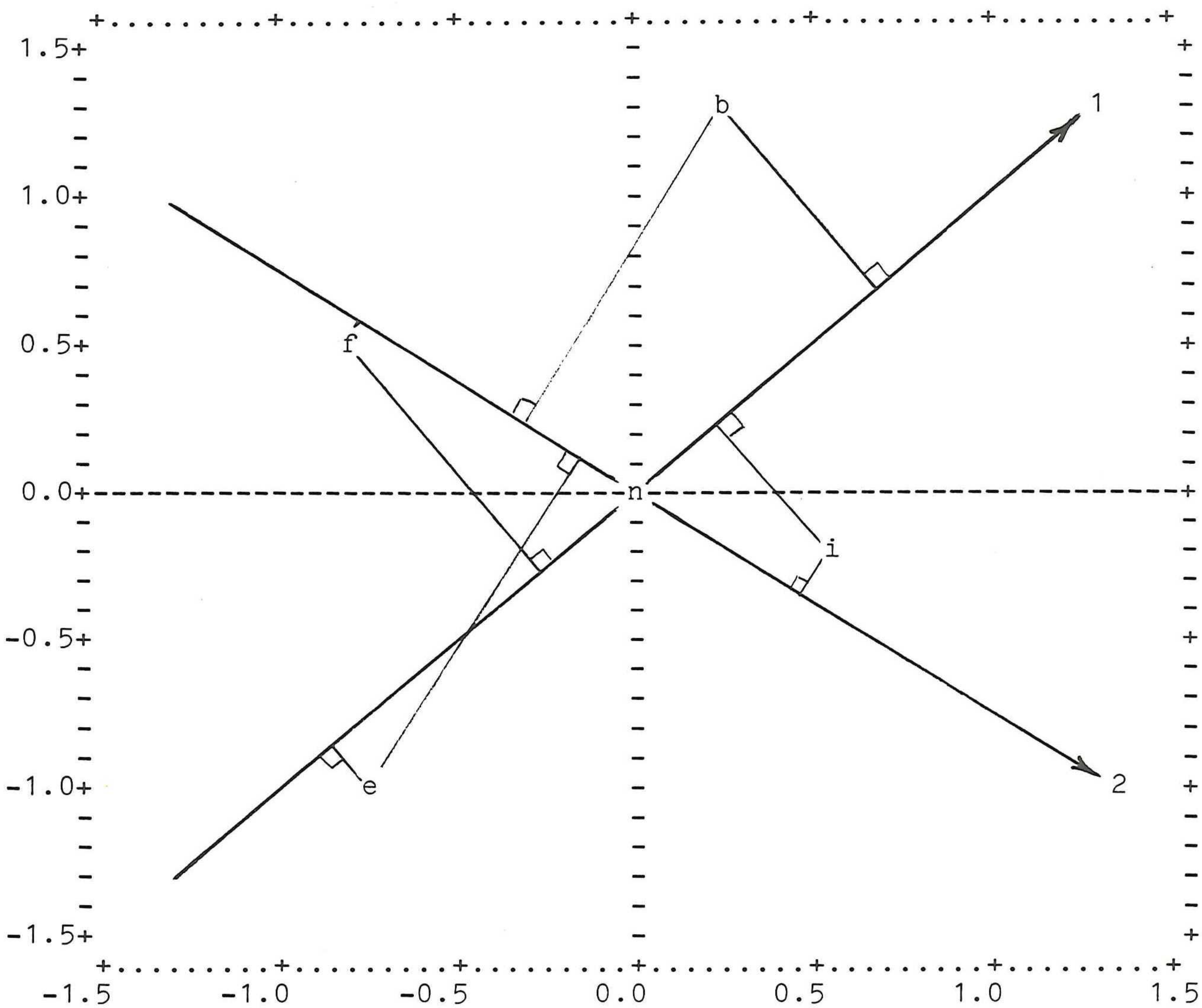


Figure 1: ILLUSTRATION OF POINT-VECTOR MODEL

THE SIX POINTS PROJECT ON VECTOR 1: e f n i b
THE SIX POINTS PROJECT ON VECTOR 2: f b e n i

Our earlier analyses using MDPREF produced a solution in 2-dimensional space, and the input matrix was collapsed across sentences for one analysis, and across speakers for a second analysis. Considerable structure in the data was lost by collapsing the input matrix, and as a result about 90% of the variance in the data was accounted for by the first dimension in the two-dimensional analyses. Even so, the second dimension proved to be easily interpretable, on the basis of the known parameters of the systems, and of the characteristics of the sentences, and the talkers who read them. In addition, the two analyses (one of which looked for the effects of different sentence material, and the other of different talkers) assigned different values of the second dimension to the fourteen systems. This suggests that the two second dimensions may be at least partially independent, and that a three-dimensional analysis might confirm this.

There are two reasons why a three-dimensional analysis is not appropriate on the collapsed data. One is that when 90% of the variance is accounted for by the first dimension, the remaining dimensions are not very reliable, since they tend to be obscured by noise (random variability) in the data. In fact the percentage of the variance accounted for by each of the first four dimensions was 1: 89.7%, 2: 4.9%, 3: 1.6% and 4: 1.0% for the data collapsed across speakers, and 1: 91.8%, 2: 6.7%, 3: 1.9% and 4: 1.0% for the data collapsed across sentences. In both cases, it is marginally meaningful to attempt to interpret the second dimensions, and meaningless for the third and fourth. Second, if the second

dimensions in the foregoing analyses were in fact independent, the axes corresponding to them in the N-dimensional space would be orthogonal. Thus collapsing the data along either dimension removes exactly that variability that might have been explained by the third dimension. Therefore, a three-dimensional MDPREF analysis was performed on the uncollapsed input data (the full 36 x 14 matrix), for both the rating and for the ranking data.

B. Rating Results

We will consider the results of analyzing the rating data fist. MDPREF produces the results of its analysis in two forms: 1) tables giving the coordinates assigned to the system points and to the sentence-speaker vectors, together with various intermediate results and other data, and 2) plots of the positions of the points and vectors for all pairs of the 3 dimensions (since 3-dimensional plots are unwieldy).

The first dimension accounted for 70.4% of the total variance in the data, and the second and third dimensions accounted for 8.9% and 6.0% respectively. Preliminary inspection of the positions of the system points in the plot of dimension 2 vs. dimension 3 suggested very strongly that not only were the effects of sentences and speakers obtained in the 2-dimensional analysis present, but they were also orthogonal - i.e. totally independent. However, the factors were inclined to the axes at an angle of about 35 degrees. Since the orientation of the axes is arbitrary in an MDPREF analysis, a program was written to perform rotations of the axes, and produce plots of the rotated points and vectors. One rotation was made about each axis: first about the major axis, so as to align the speaker and sentence factors with axes 2 and 3, and then two minor rotations were performed to bring the point representing undistorted speech onto the main axis (i.e. to give it zero coordinates on dimensions 2 and 3). The distribution of the points in the space, after rotation, are given in Figs. 2 through 5.

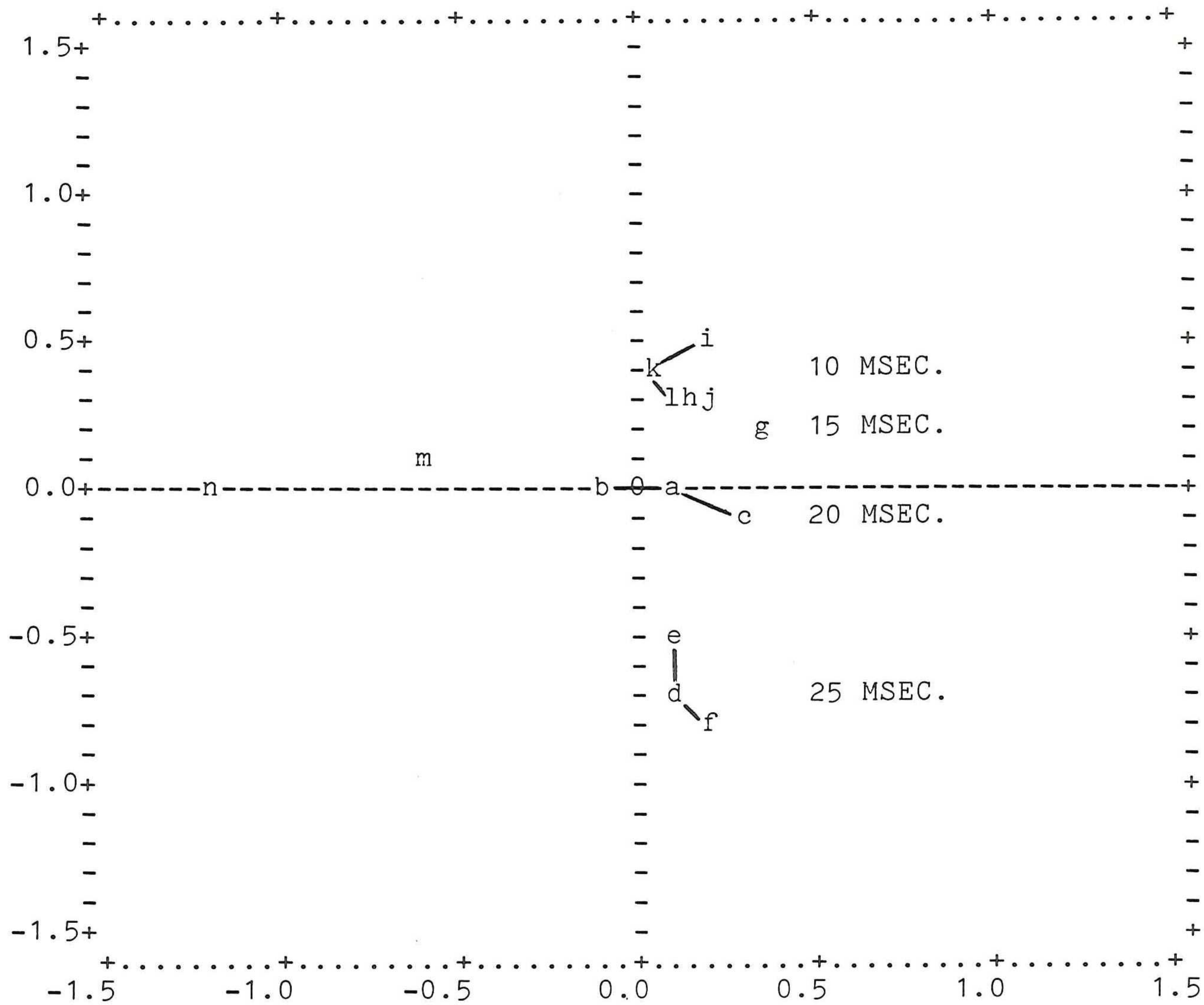


Figure 2: PLOT OF VOCODER SYSTEMS = POINTS: DIM 1 (X) VS DIM 2 (Y)

SYSTEM	# POLE	FRAME SIZE	VAR-RATE THRESH	STEP SIZE	EXPECT BITS PER	FOUND SECOND
A	12	20		1.0 DB	2650	2630
B	10	20		0.6	2650	2633
C	14	20		1.4	2700	2681
D	12	25		0.45	2640	2610
E	14	25		0.7	2640	2612
F	10	25		0.2	2680	2652
G	10	15		1.75	2666	2618
H	12	10	1.5 DB	0.5	2660	2574
I	12	10	1.0	1.0	2650	2652
J	12	10	1.75	0.25	2627	2687
K	14	10	1.5	0.6	2685	2766
L	12	15	1.5	0.4	2600	2535
M	14	10	VOCODED, BUT UNQUANTIZED			
N	ORIGINAL WAVEFORM, DIGITIZED AND RECONSTITUTED (IE PCM)					

Consider Fig. 2 first. The X-axis is about 23 degrees from the major dimension as produced by the analysis, and so probably accounts for close to 70% of the variance. This main dimension corresponds closely to the main dimension obtained earlier from the two-dimensional analyses performed on the collapsed data. The positions of the systems on the x-axis fall into three groups. The left-most point (system n) corresponds to the original speech. The next point to the right (system m) corresponds to the vocoded but unquantized speech. The rest of the systems fall in a fairly small group, with a few systems on the edges, such as system b at the left and g at the right. (Remember we are only considering the distribution along the x-axis at present.)

The major axis can probably be interpreted as "overall clarity" of the speech quality obtained from a system. The original speech is much clearer than the unquantized speech which in turn is much clearer than that produced by the remaining systems. On the other hand, there are substantial differences in the amount of information needed to code the waveform, which ranges from about 90,000 bps for the PCM original speech down to about 2600 bps for the twelve vocoders. Presumably, a wider range of bit rates in the vocoders under test would have led to a wider distribution along the x-axis. On the other hand, had the original (n) and unquantized (m) systems not been included in the tests, the proportion of the variance accounted for by the present dimension 1 would have been substantially smaller.

Now let us turn to the Y-axis of Fig. 2. The second dimension performs a perfect separation of the twelve vocoders as a function of their <u>frame size</u>. (Frame sizes are indicated on the figures.) Lines join all systems having the same frame size, which varied between 10 msec, for all but one of the variable-rate systems, through 25 msec for systems d, e, and f. This separation is very similar (but much more efficient and convincing) than that obtained in the analysis of the effects of sentence material from the collapsed data (see QPR #3, Fig. 2 and note that the sense of the Y-axis is reversed). It suggests that the major difference in perceived quality, shown by the twelve vocoders, was a result of the varied ability of the systems to track rapid spectral changes.

We will postpone detailed discussion of the speaker-sentence vectors until later in the report. In general, the placing of the vectors is in good agreement with the placing obtained from the 2-dimensional analyses reported in QPR #3. The vectors for the slowly-changing Sentence 1 (Why were you away a year, Roy?) tend to move from low left to high right in Fig. 2, showing that the 25 msec frame systems (d, e, f) gave relatively high quality on this sentence, and the shorter frame systems did relatively worse. This finding can be explained as a result of our decision to equate the bit rate of all systems to 2600 bps. The slower the frame rate, the more bits are available for specifying the spectrum transmitted by a single frame, and the better the spectral match. The long frame size was never a disadvantage, since all changes were relatively slow in Sentence 1. On the other hand, the systems with shorter

frames had less bits available for specifying the spectrum, and there were no acoustic events in Sentence 1 that used the ability of these systems to follow rapid spectral changes. The relative qualities were reversed for the rapidly-changing Sentence 4 (Which tea-party did Baker go to?). Here, the long frame systems were unable to follow the rapid changes, and significantly degraded the speech. Their better ability to match the spectrum was not sufficient to offset the degradation.

These findings suggest that dimension 2 can perhaps be interpreted as "ability to follow rapid spectral changes without large errors" or "ability to model spectrum dynamics."

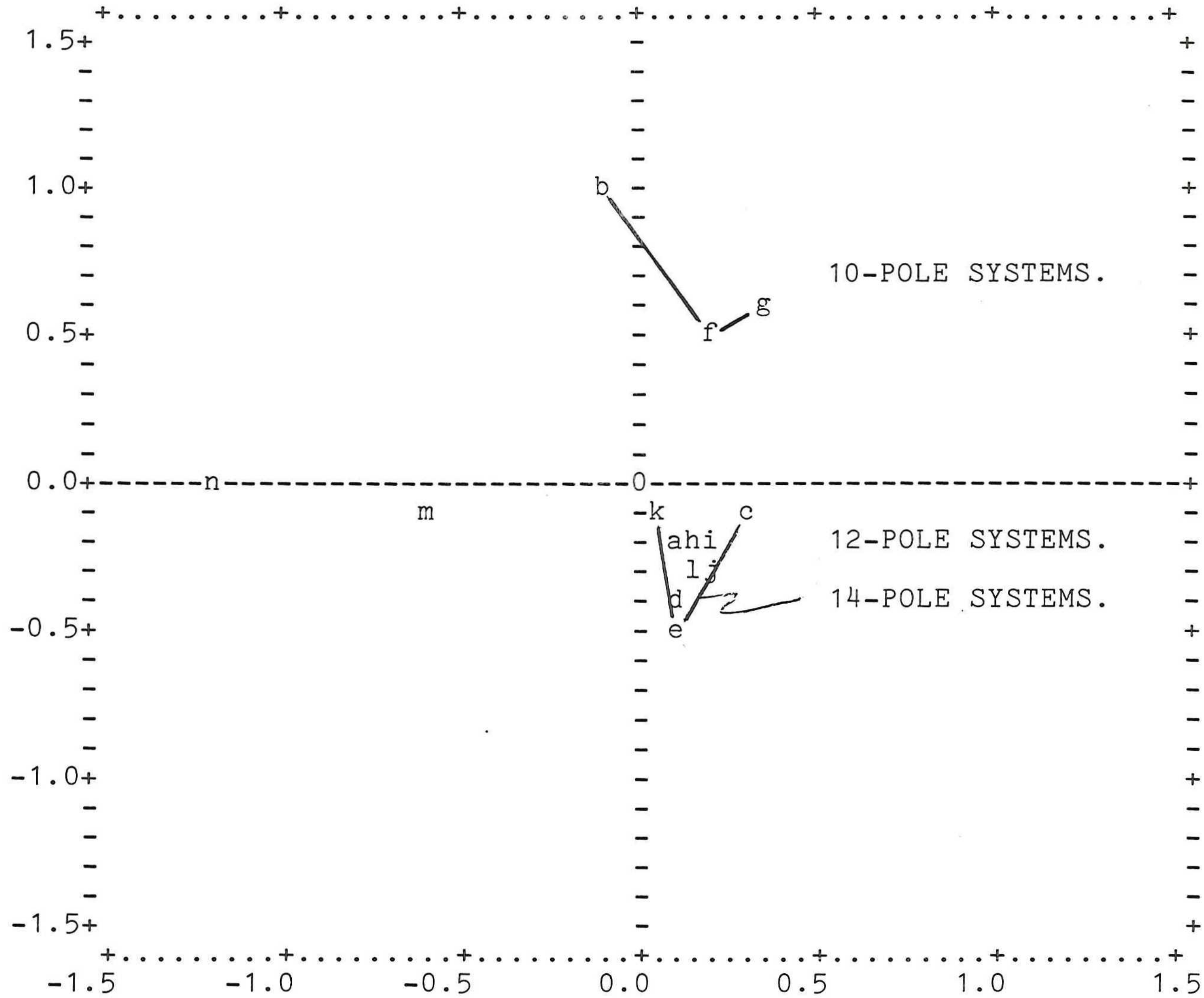


Figure 3: PLOT OF VOCODER SYSTEMS = POINTS: DIM 1 (X) VS DIM 3 (Y)

SYSTEM	# POLE	FRAME SIZE	VAR-RATE THRESH	STEP SIZE	EXPECT BITS PER SECOND	FOUND BITS PER SECOND
A	12	20		1.0 DB	2650	2630
B	10	20		0.6	2650	2633
C	14	20		1.4	2700	2681
D	12	25		0.45	2640	2610
E	14	25		0.7	2640	2612
F	10	25		0.2	2680	2652
G	10	15		1.75	2666	2618
H	12	10	1.5 DB	0.5	2660	2574
I	12	10	1.0	1.0	2650	2652
J	12	10	1.75	0.25	2627	2687
K	14	10	1.5	0.6	2685	2766
L	12	15	1.5	0.4	2600	2535
M	14	10	VOCODED, BUT UNQUANTIZED			
N	ORIGINAL WAVEFORM, DIGITIZED AND RECONSTITUTED (IE PCM)					

Figure 3 presents a second view of the placing of the 14 systems in the three dimensional solution space. The positions of the systems on the third dimension are plotted against those on the first dimension. The distribution of points on the X-axis is the same as in Fig. 2, but now the Y-axis separates the systems by the number of poles used in modelling the spectrum. The separation corresponds to the separation achieved in the two-dimensional analysis on the collapsed data, as reported in QPR #3, Fig. 3 (note that the <u>sense</u> of the Y-axis is reversed), except that the separation is more convincing. To be more accurate, Dimension 3 separates the systems that used only 10 poles from those using 12 or 14. Systems with 12 poles were not well separated from those with 14. In the earlier analysis, (QPR #3), this separation was achieved as a result of the varied properties of the different speakers voices.

Although we will postpone detailed discussion of the placing of the vectors in the present 3-dimensional analysis until later in this report, the separation here also was achieved as a result of the different properties of the speakers voices, especially their fundamental frequency. As we shall see, five of the six vectors for the talker with the lowest fundamental (Speaker #3, DK) run from lower left to upper right in Fig. 3, whereas those for speakers with the highest fundamentals run nearly parallel with the X-axis.

Figure 3 shows that the 10-pole systems produced substantial degradation of speech quality when processing sentences by Speaker

#3. Informal listening shows that these sentences sound muffled, or as if low-pass filtered. There are two possible explanations for this, and it is not clear which is correct. A speaker with a low fundamental tends to be a large man, and to have a long vocal tract. As a result, all formant frequencies are lower than the population average, and the 5 KHz pass band of the vocoder systems may include more high-order formants for these speakers. The 10-pole systems do not have enough poles available to match the spectra having the extra formants and the failure is most apparent in the higher frequencies. The second possible explanation is that a low fundamental effectively specifies the spectrum in more detail, since there are many more harmonics of the fundamental in the pass-band, and the amplitude of each harmonic specifies the spectral level at that frequency. As a result, more poles are needed to produce an adequate spectral match at low frequency, and the 10-pole systems have too few remaining poles to model adequately the higher frequencies. Alternatively, the high-order formants may fail to be separated by the widely separated harmonics of a high fundamental, so they can be matched by very few poles, whereas the formants <u>are</u> separated by the closely spaced harmonics of a low fundamental, and require more poles for an adequate match.

From the foregoing arguments, it seems likely that Dimension 3 can be interpreted as reflecting the best match that can be obtained between the observed spectrum and the modelled spectrum, or the "best static spectral match," as contrasted with the "lowest dynamic spectral error" for Dimension 2.

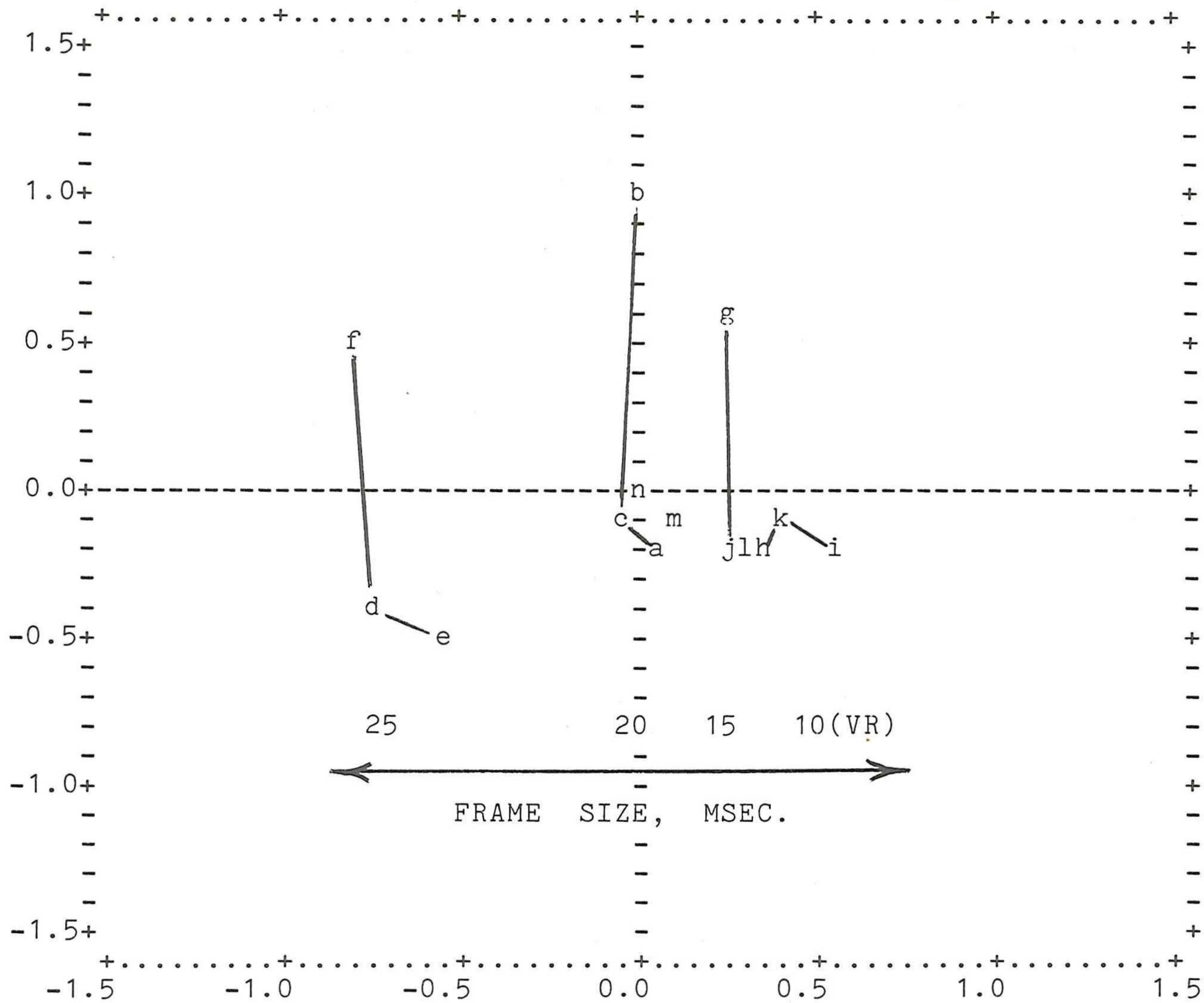


Figure 4: PLOT OF VOCODER SYSTEMS = POINTS: DIM 2 (X) VS DIM 3 (Y)

SYSTEM	# POLE	FRAME SIZE	VAR-RATE THRESH	STEP SIZE	EXPECT BITS PER SECOND	FOUND
A	12	20		1.0 DB	2650	2630
B	10	20		0.6	2650	2633
C	14	20		1.4	2700	2681
D	12	25		0.45	2640	2610
E	14	25		0.7	2640	2612
F	10	25		0.2	2680	2652
G	10	15		1.75	2666	2618
H	12	10	1.5 DB	0.5	2660	2574
I	12	10	1.0	1.0	2650	2652
J	12	10	1.75	0.25	2627	2687
K	14	10	1.5	0.6	2685	2766
L	12	15	1.5	0.4	2600	2535
M	14	10	VOCODED, BUT UNQUANTIZED			
N	ORIGINAL WAVEFORM, DIGITIZED AND RECONSTITUTED (IE PCM)					

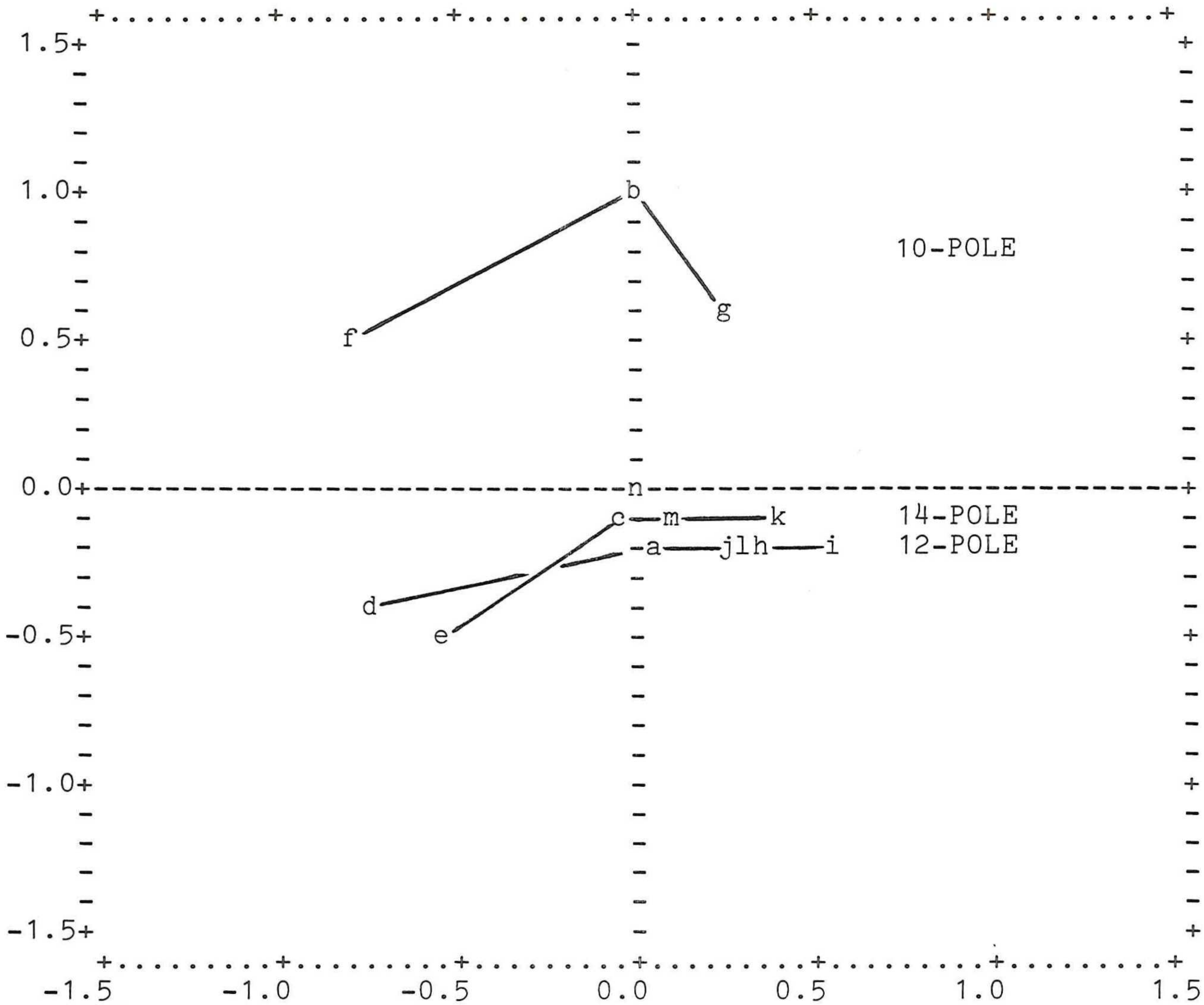


Figure 5: PLOT OF VOCODER SYSTEMS = POINTS: DIM 2 (X) VS DIM 3 (Y)

SYSTEM	# POLE	FRAME SIZE	VAR-RATE THRESH	STEP SIZE	EXPECT BITS PER SECOND	FOUND
A	12	20		1.0 DB	2650	2630
B	10	20		0.6	2650	2633
C	14	20		1.4	2700	2681
D	12	25		0.45	2640	2610
E	14	25		0.7	2640	2612
F	10	25		0.2	2680	2652
G	10	15		1.75	2666	2618
H	12	10	1.5 DB	0.5	2660	2574
I	12	10	1.0	1.0	2650	2652
J	12	10	1.75	0.25	2627	2687
K	14	10	1.5	0.6	2685	2766
L	12	15	1.5	0.4	2600	2535
M	14	10	VOCODED, BUT UNQUANTIZED			
N	ORIGINAL WAVEFORM, DIGITIZED AND RECONSTITUTED (IE PCM)					

Figures 4 and 5 show the third projection of the 14 system points, now with Dimension 2 plotted against Dimension 3. In this view, the vectors point towards the reader, out of the page. In Fig. 4, lines are drawn joining the systems with the same frame size, and in Fig. 5 the lines join the systems using the same number of poles in their spectral modelling. Comparison of Figs. 4 and 5 shows that these two separations are almost exactly orthogonal, indicating that their effects on perceived quality are <u>independent</u>. This is a very important finding, since it suggests that the two variables can be optimized <u>independently</u>.

Now let us consider in more detail the distribution of the 36 vectors, one for each of the speaker-sentence combinations.

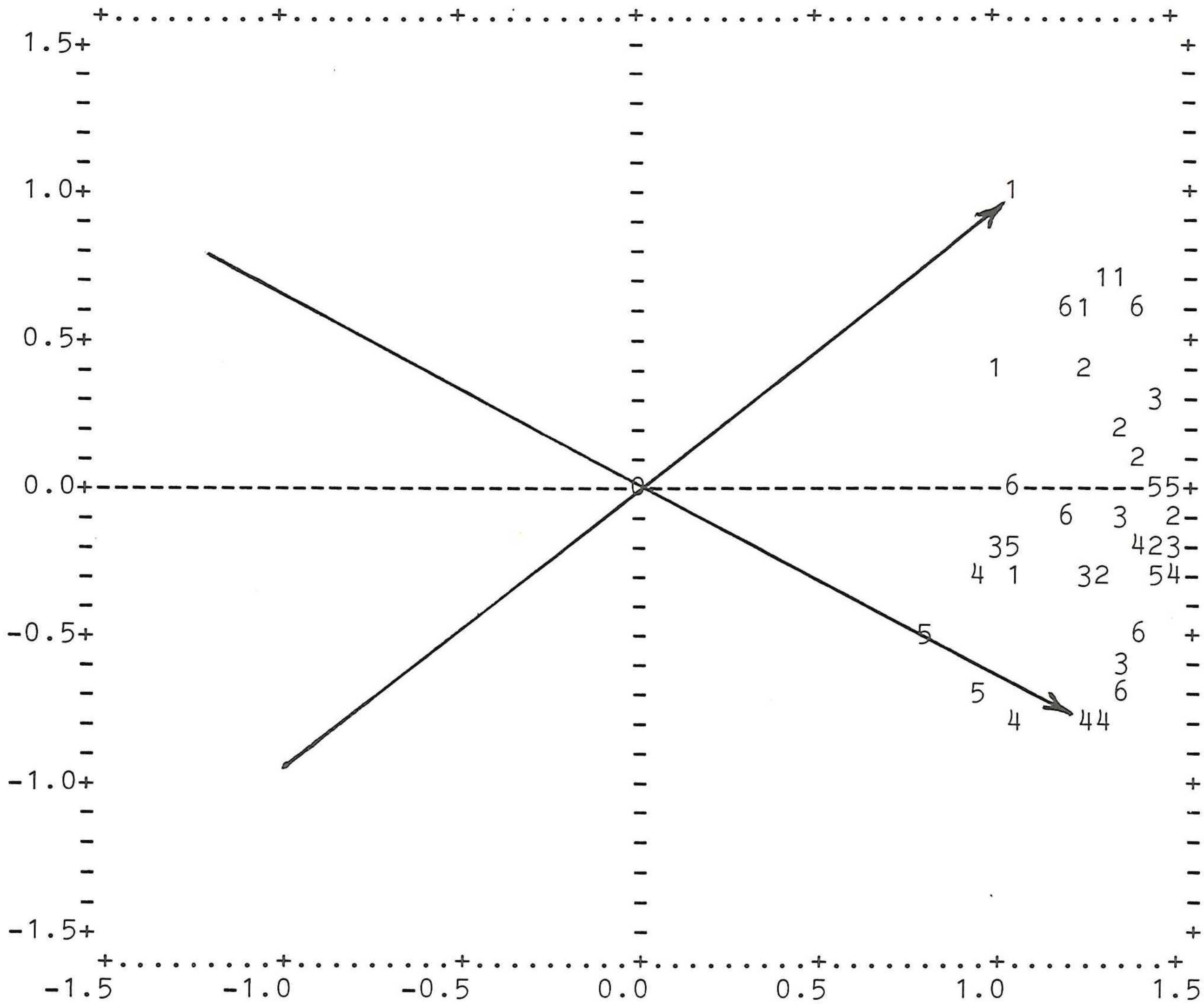


Figure 6: PLOT OF SPEAKER*SENTENCE VECTORS: DIM 1 (X) VS DIM 2 (Y)

PLOTTED DIGITS REPRESENT ENDS OF SENTENCE VECTORS
VECTOR 1 = SPEAKER 1 (AR), SENTENCE 1
VECTOR 4 = SPEAKER 1 (AR), SENTENCE 4

SENTENCES

1. WHY WERE YOU AWAY A YEAR, ROY?
2. NANNY MAY KNOW MY MEANING.
3. HIS VICIOUS FATHER HAS SEIZURES.
4. WHICH TEA-PARTY DID BAKER GO TO?
5. THE LITTLE BLANKETS LAY AROUND ON THE FLOOR.
6. THE TROUBLE WITH SWIMMING IS THAT YOU CAN DROWN.

Figure 6 plots the positions of the ends of the vectors, and the number labelling each vector corresponds to the number of the sentence. (The data in this figure and those in Fig. 2 are combined in Fig. 1 of the appendix, where the digits and 26 capital letters are used to label the vectors for both speaker and sentence, and the 14 systems are labelled with lower case letters a-n.) It can be seen that the vectors for the slowly varying Sentence 1 tend to be high on the plot, and those for the rapidly changing Sentence 4 tend to be at the bottom.

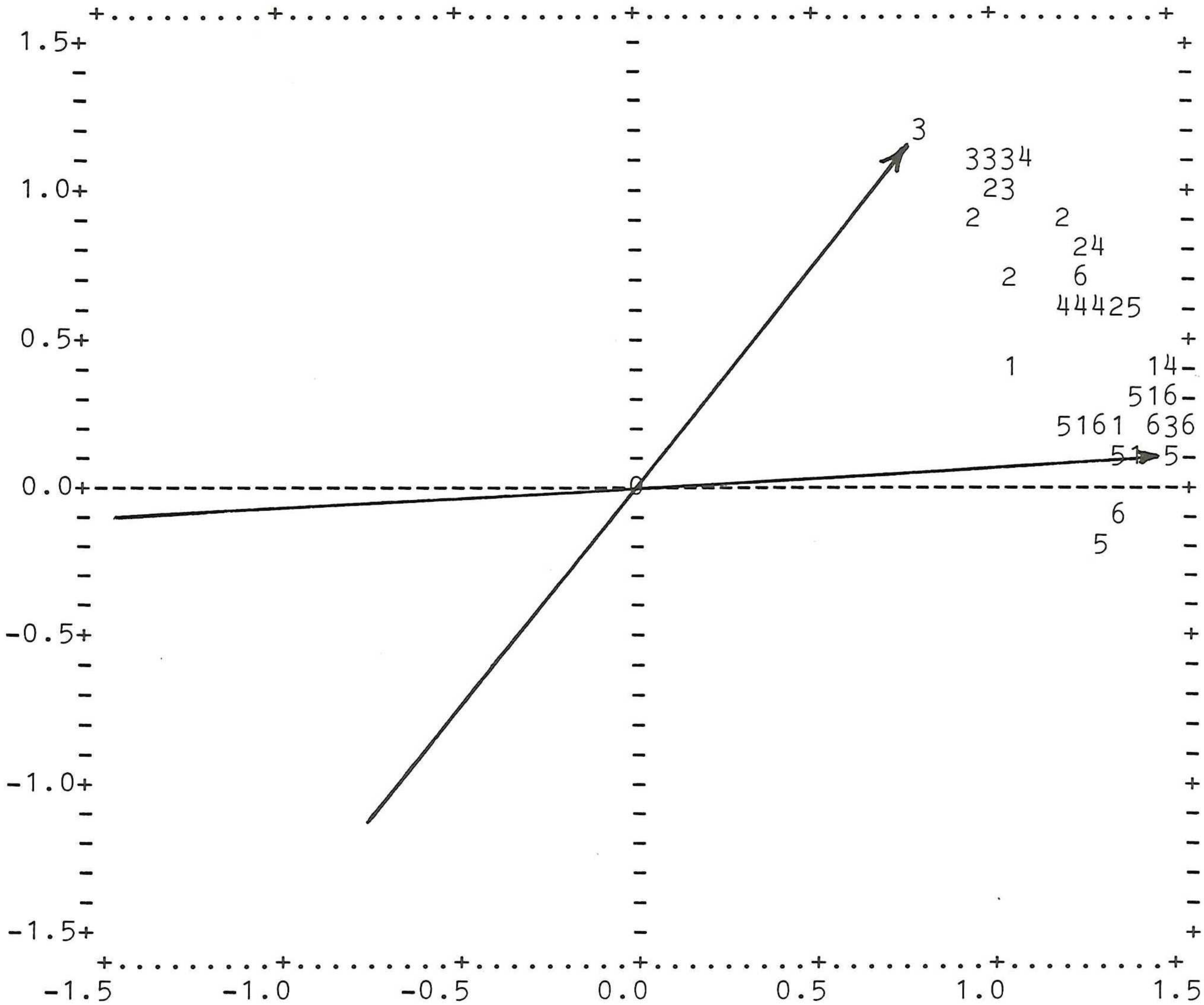


Figure 7: PLOT OF SPEAKER*SENTENCE VECTORS: DIM 1 (X) VS DIM 3 (Y)

PLOTTED DIGITS REPRESENT ENDS OF SPEAKER VECTORS
VECTOR 3 = SPEAKER 3 (DK), SENTENCE 5
VECTOR 5 = SPEAKER 5 (RS), SENTENCE 5

TALKERS

NO	SEX	ID	AV-FO	NASALITY	DURATION
1	F	AR	167 HZ	15.2 DB	2.00 SECS
2	M	JB	118	14.3	2.20
3	M	DK	95	17.3	2.05
4	M	DD	139	16.4	2.00
5	F	RS	209	17.0	2.40
6	F	PF	232	17.5	2.70

Figure 7 shows the dimension 1 vs dimension 3 plot of the end points of the vectors for the 36 speaker-sentence combinations. The labelling digits in this plot refer to the speakers. (The data in this figure and in Fig. 3 are combined in Fig. 2 of the Appendix, where the labelling permits identification of both speaker and sentence.) It can be seen that all but one of the vectors associated with speaker 3 cluster towards the top of the figure, followed by those for speakers 2 and 4, then 1, 5, and 6. The separations of vectors by sentence in Fig. 6, and by speaker in Fig. 7, are both similar to the separations obtained in the earlier analyses on the collapsed data (QPR #3). In addition, the wider separation obtained in the present analysis suggests that substantially more of the variance in the input data is accounted for than in the earlier analysis. Inspection of Figs. 6 and 7 shows that the separation of the vectors by sentence and speaker by dimensions 2 and 3 is not perfect, and we plan to look for other factors that might explain this.

C. Comparison of Rating and Ranking Data

The success of the three-dimensional analysis of the rating data suggested that a similar analysis of the ranking data might yield a powerful test of the idea that all quality judgments are made on the basis of a single underlying psychological structure. If the two analyses agree in how they place the vocoder systems in the solution space, this would be strong evidence that the same structure is present in both sets of data, and would allow the less

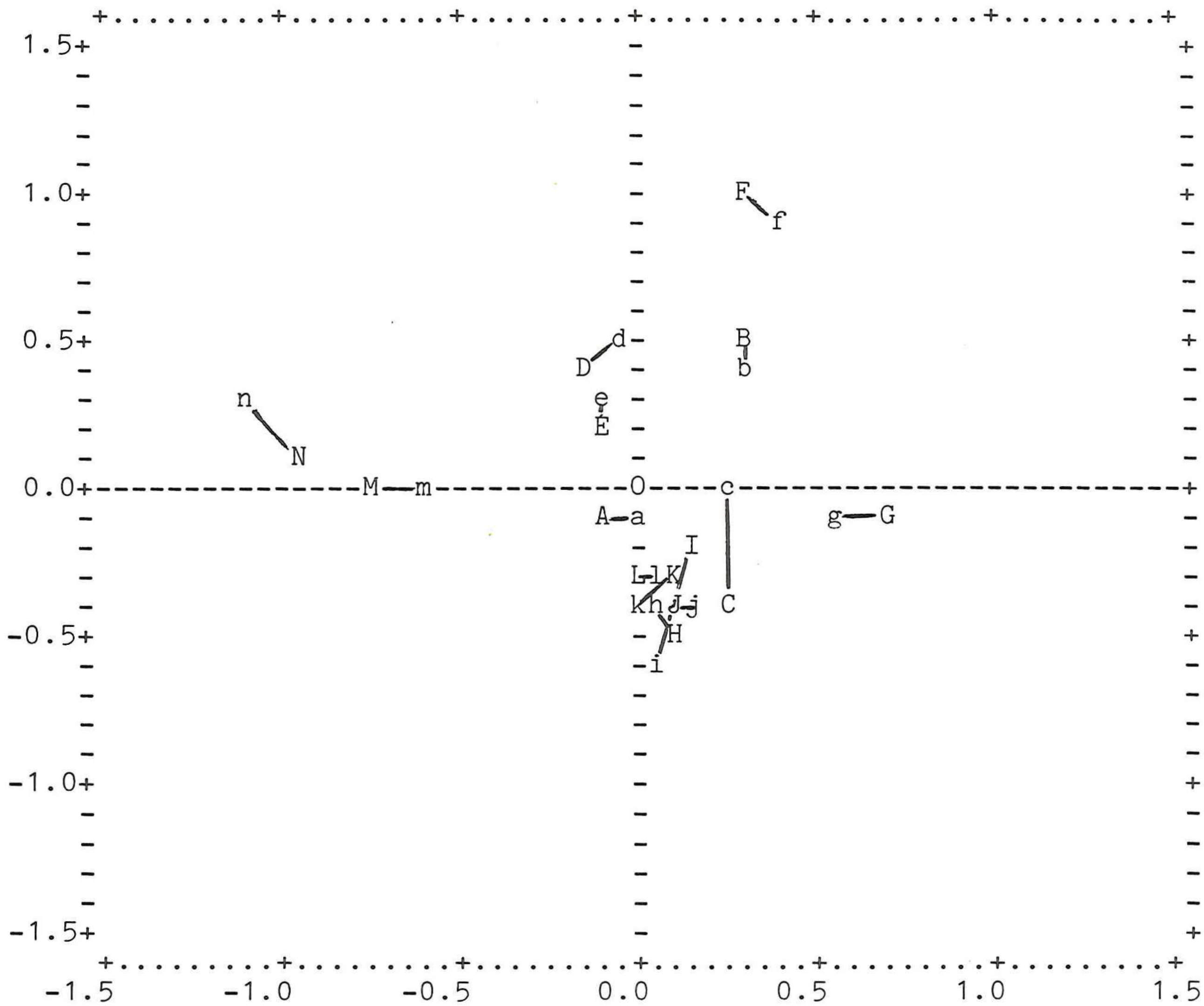


Figure 8: COMPARISON OF PLACING OF SYSTEM-POINTS IN SOLUTION SPACE, FROM RATING DATA (LOWER-CASE) AND FROM RANKING DATA (UPPER-CASE) LABELS. DIMENSION 1 (X) VS. DIMENSION 2 (Y).

SYSTEM	# POLE	FRAME SIZE	VAR-RATE THRESH	STEP SIZE	EXPECT BITS PER SECOND	FOUND
A	12	20		1.0 DB	2650	2630
B	10	20		0.6	2650	2633
C	14	20		1.4	2700	2681
D	12	25		0.45	2640	2610
E	14	25		0.7	2640	2612
F	10	25		0.2	2680	2652
G	10	15		1.75	2666	2618
H	12	10	1.5 DB	0.5	2660	2574
I	12	10	1.0	1.0	2650	2652
J	12	10	1.75	0.25	2627	2687
K	14	10	1.5	0.6	2685	2766
L	12	15	1.5	0.4	2600	2535
M	14	10	VOCODED, BUT UNQUANTIZED			
N	ORIGINAL WAVEFORM, DIGITIZED AND RECONSTITUTED (IE PCM)					

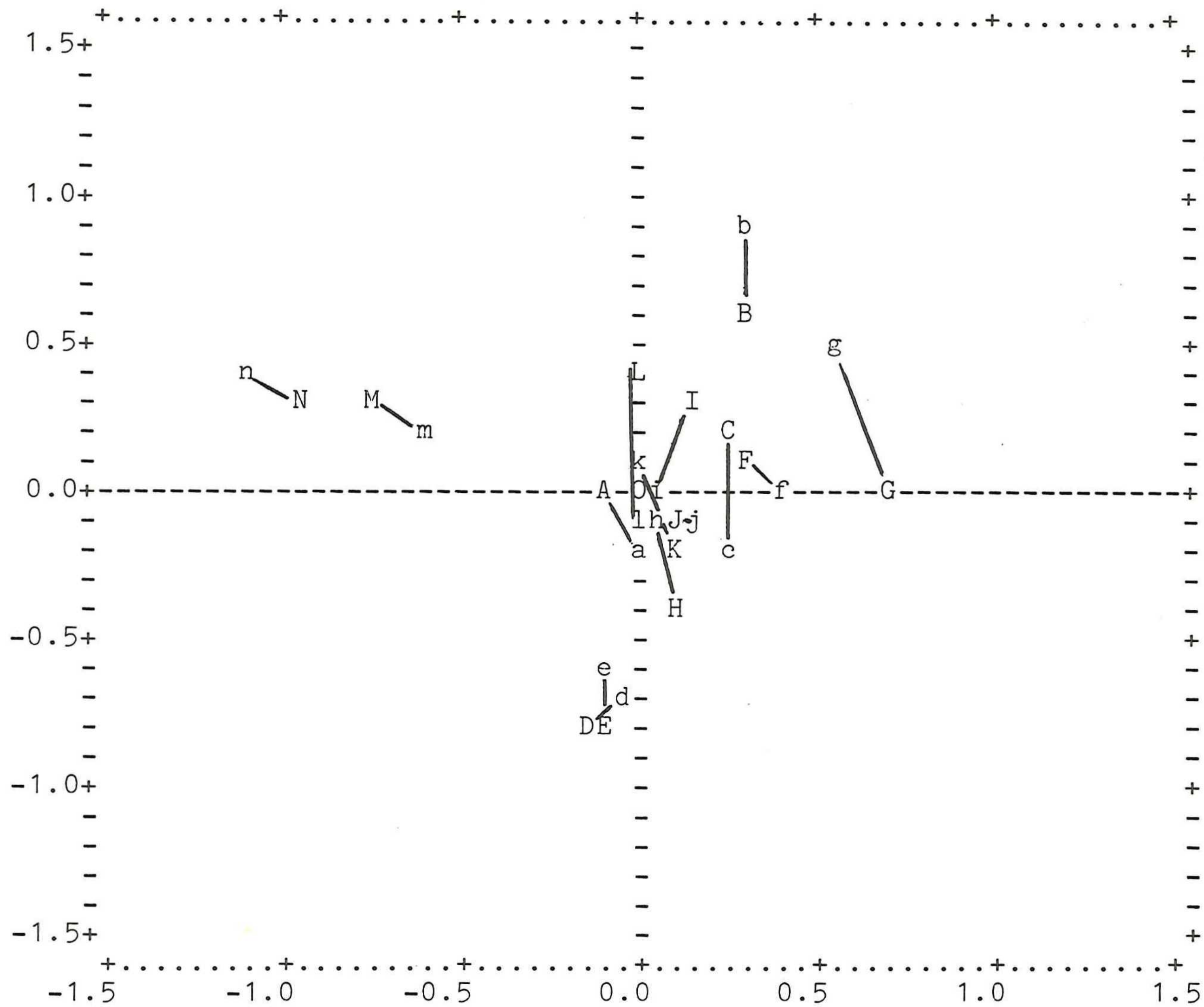


Figure 9: COMPARISON OF PLACING OF SYSTEM-POINTS IN SOLUTION SPACE, FROM RATING DATA (LOWER-CASE) AND FROM RANKING DATA (UPPER-CASE) LABELS. DIMENSION 1 (X) VS. DIMENSION 3 (Y).

SYSTEM	# POLE	FRAME SIZE	VAR-RATE THRESH	STEP SIZE	EXPECT BITS PER	FOUND SECOND
A	12	20		1.0 DB	2650	2630
B	10	20		0.6	2650	2633
C	14	20		1.4	2700	2681
D	12	25		0.45	2640	2610
E	14	25		0.7	2640	2612
F	10	25		0.2	2680	2652
G	10	15		1.75	2666	2618
H	12	10	1.5 DB	0.5	2660	2574
I	12	10	1.0	1.0	2650	2652
J	12	10	1.75	0.25	2627	2687
K	14	10	1.5	0.6	2685	2766
L	12	15	1.5	0.4	2600	2535
M	14	10	VOCODED, BUT UNQUANTIZED			
N	ORIGINAL WAVEFORM, DIGITIZED AND RECONSTITUTED (IE PCM)					

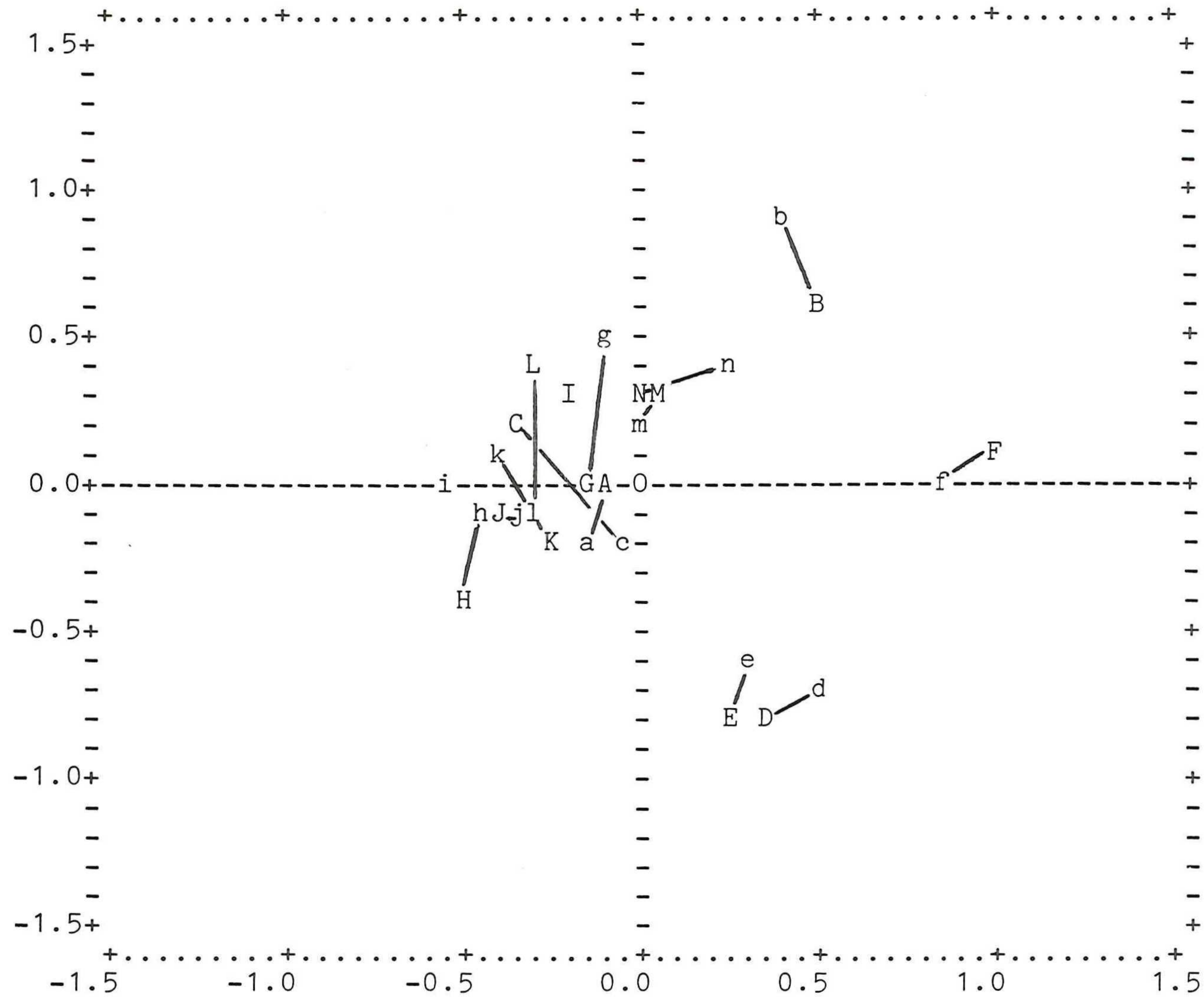


Figure 10: COMPARISON OF PLACING OF SYSTEM-POINTS IN SOLUTION SPACE, FROM RATING DATA (LOWER-CASE) AND FROM RANKING DATA (UPPER-CASE) LABELS. DIMENSION 2 (X) VS. DIMENSION 3 (Y).

SYSTEM	# POLE	FRAME SIZE	VAR-RATE THRESH	STEP SIZE	EXPECT BITS PER SECOND	FOUND BITS PER SECOND
A	12	20		1.0 DB	2650	2630
B	10	20		0.6	2650	2633
C	14	20		1.4	2700	2681
D	12	25		0.45	2640	2610
E	14	25		0.7	2640	2612
F	10	25		0.2	2680	2652
G	10	15		1.75	2666	2618
H	12	10	1.5 DB	0.5	2660	2574
I	12	10	1.0	1.0	2650	2652
J	12	10	1.75	0.25	2627	2687
K	14	10	1.5	0.6	2685	2766
L	12	15	1.5	0.4	2600	2535
M	14	10	VOCODED, BUT UNQUANTIZED			
N	ORIGINAL WAVEFORM, DIGITIZED AND RECONSTITUTED (IE PCM)					

efficient procedure to be abandoned for further testing. Figures 8, 9 and 10 show the three two-dimensional views of the 14 vocoder systems, for both the rating solution (labelled with lower-case letters) and the ranking solution (labelled with upper case letters). Lines join the placing of each system by the rating analysis to that by the ranking analysis. The shortness of the lines attests to the very close overall agreement between the two solutions. The correlation between the first dimension coordinates of the systems in the rating solution with those in the ranking solution, is 0.973 - an almost perfect correlation. On the second dimension, the coefficient is 0.916, and on the third, 0.755. The first two of these are significant well beyond the 0.001 level, and the third is significant at the 0.01 level. Neither solution was rotated before the results were plotted: it is possible that the agreement could be made even better by minimizing the least-squares of the distances, but the agreement is already good enough to prove our point. The fact that the ranking and rating tasks yield such similar solutions means that the more efficient procedure can be used exclusively in future testing. We plan to give the rating task to a larger group of naive subjects, to confirm the generality of the result, and it may also be appropriate to compare it with the results obtained by other measurement methods such as the PAR test (Voiers, forthcoming).

III. ANALYSES WITH INDSCAL

We have abandoned further attempts to analyze our rating and ranking data with INDSCAL (a second multidimensional scaling program we obtained from Bell Telephone Laboratories). The major reason is that despite repeated attempts, the INDSCAL analyses resisted our efforts at interpretation. There is no doubt that the INDSCAL model is more powerful than the MDPREF model, in that it requires fewer assumptions about the psychological structures being modelled. On the other hand, to be fully appropriate, INDSCAL requires its input data to be in the form of paired comparisons, which was impracticable with the large number of stimuli available in the present experiment. Only the canonical decomposition part of INDSCAL is appropriate for preference data such as ours. It may be that consultation with the developer of the model (J.D. Carroll at Bell Labs) might resolve our difficulties, but we have not pursued this course, largely because of the considerable success we have had with MDPREF.

IV. PHONEME SPECIFIC TESTS

We have performed a pilot run of the Phoneme Specific Test described in earlier progress reports. We selected seven lists, and processed each list through five vocoder systems, including PCM (the "Original Waveform" vocoder in our other tests). Twenty subjects served. We have not analyzed the results in detail, for two reasons. First, it became clear very quickly that subjects made

very few errors, and therefore a more detailed response was needed to improve the efficiency and usefulness of the test. Secondly, we found that the learning effects, as subjects were exposed to each of the seven lists five times, were substantially larger than expected. We are in the process of preparing for a repeat of the experiment, with a modified procedure. We will use a secondary task (writing down a digital clock reading after responding to each speech stimulus) to give us information about how quickly the subject arrived at his decision. This will permit us to gain useful information from the correct responses as well as from the errors, but will not interfere with the confusion matrices which were the original object of the test.

V. FUTURE PLANS

1. We will run the rating test on a larger sample of naive subjects, to confirm the validity of our earlier conclusions.

2. We will explore in more detail the objective parameters of each speaker's production of each sentence, and try to account for the placing of the vectors in the solution space. The results of this further analysis should be of direct use in developing objective quality evaluation procedures.

3. We will perform an extended trial of the phoneme specific tests. The viability of this test would be greatly enhanced by the availability of real-time implementation of the vocoder systems under test, and this is an approach we may explore.

4. We are preparing a larger set of Phoneme Specific <u>sentences</u> (similar to the six sentences used in our rating and ranking tasks), which we will record under conditions appropriate to real-life use of the vocoding systems. We will make these recordings available to other ARPA contractors, for them to process through their real-time systems and return to us for further evaluation tests.

APPENDIX

SENTENCES

1. WHY WERE YOU AWAY A YEAR, ROY?
2. NANNY MAY KNOW MY MEANING.
3. HIS VICIOUS FATHER HAS SEIZURES.
4. WHICH TEA-PARTY DID BAKER GO TO?
5. THE LITTLE BLANKETS LAY AROUND ON THE FLOOR.
6. THE TROUBLE WITH SWIMMING IS THAT YOU CAN DROWN.

TALKERS

NO	SEX	ID	AV-FO	NASALITY	DURATION
1	F	AR	167 Hz	15.2 dB	2.00 SECS
2	M	JB	118	14.3	2.20
3	M	DK	95	17.3	2.05
4	M	DD	139	16.4	2.00
5	F	RS	209	17.0	2.40
6	F	PF	232	17.5	2.70

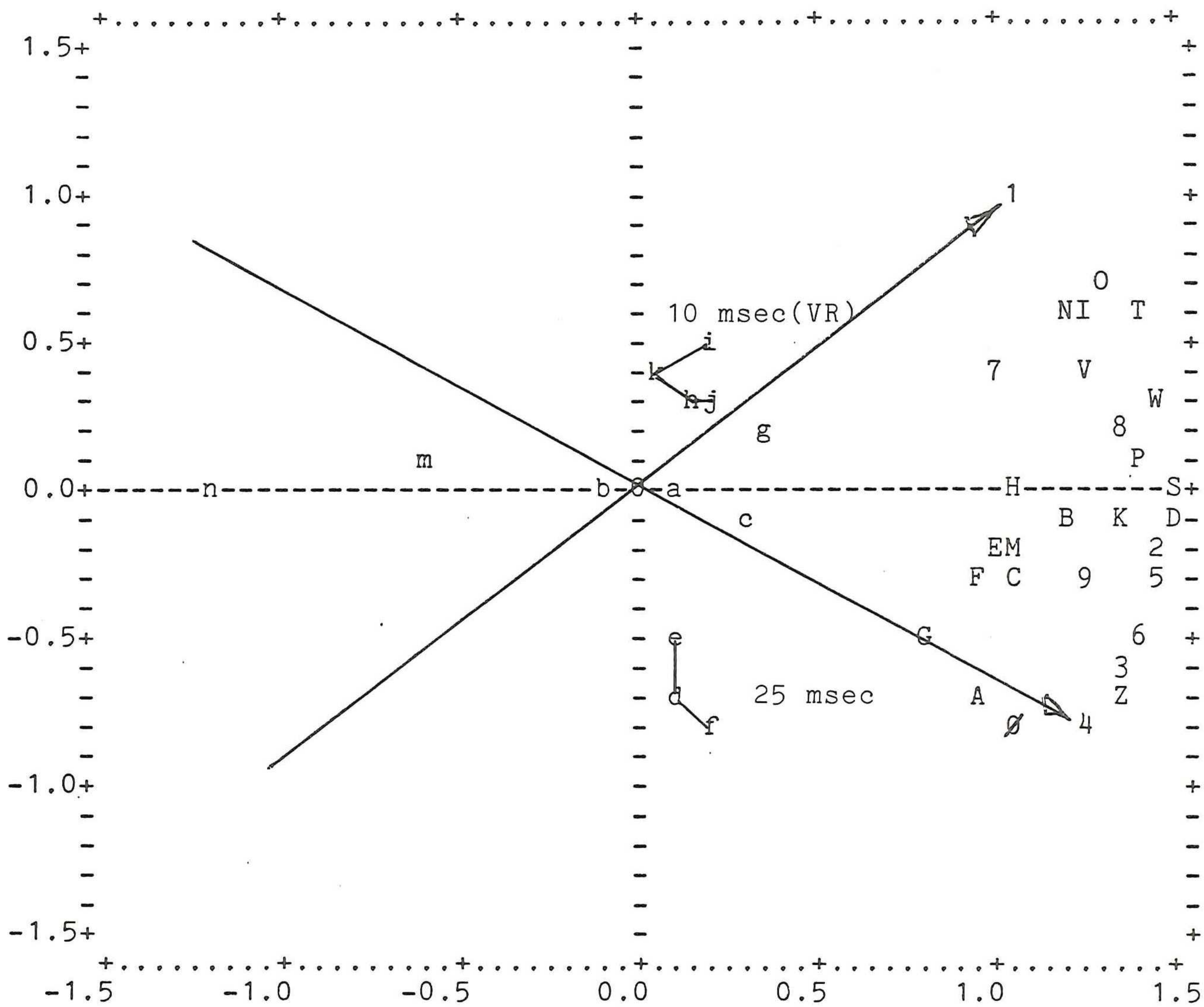
SYSTEM	# POLE	FRAME SIZE	VAR-RATE THRESH	STEP SIZE	EXPECT BITS PER SECOND	FOUND BITS PER SECOND
A	12	20		1.0 dB	2650	2630
B	10	20		0.6	2650	2633
C	14	20		1.4	2700	2681
D	12	25		0.45	2640	2610
E	14	25		0.7	2640	2612
F	10	25		0.2	2680	2652
G	10	15		1.75	2666	2618
H	12	10	1.5 dB	0.5	2660	2574
I	12	10	1.0	1.0	2650	2652
J	12	10	1.75	0.25	2627	2687
K	14	10	1.5	0.6	2685	2766
L	12	15	1.5	0.4	2600	2535
M	14	10	VOCODED, BUT UNQUANTIZED			
N	ORIGINAL WAVEFORM, DIGITIZED AND RECONSTITUTED (IE PCM)					

	SENTENCE NUMBER 1	2	3	4	5	6
SPEAKER 1=AR	1	2	3	4	5	6
2=JB	7	8	9	0	A	B
3=DK	C	D	E	F	G	H
4=DD	I	J	K	L	M	N
5=RS	O	P	Q	R	S	T
6=PF	U	V	W	X	Y	Z

TABLE 1: KEY TO SENTENCES, SPEAKER CHARACTERISTICS, VOCODER SYSTEM CHARACTERISTICS, AND LABELS ON SPEAKER*SENTENCE VECTORS.

MDPREF ANALYSIS OF DATA FILE: RTS3
REPLOT RATING DATA, 4S,36V,14P,3D; ROTATED PT2,3,2;PT14,1,2;PT14

PLOT OF SYSTEMS(PTS) AND SPKR*SENT VECTORS: DIM 1 VS DIM 2



DUPLICATE POINTS:

J=9, L=2, Q=2, R=4, U=0, X=5, Y=S, 1=h,

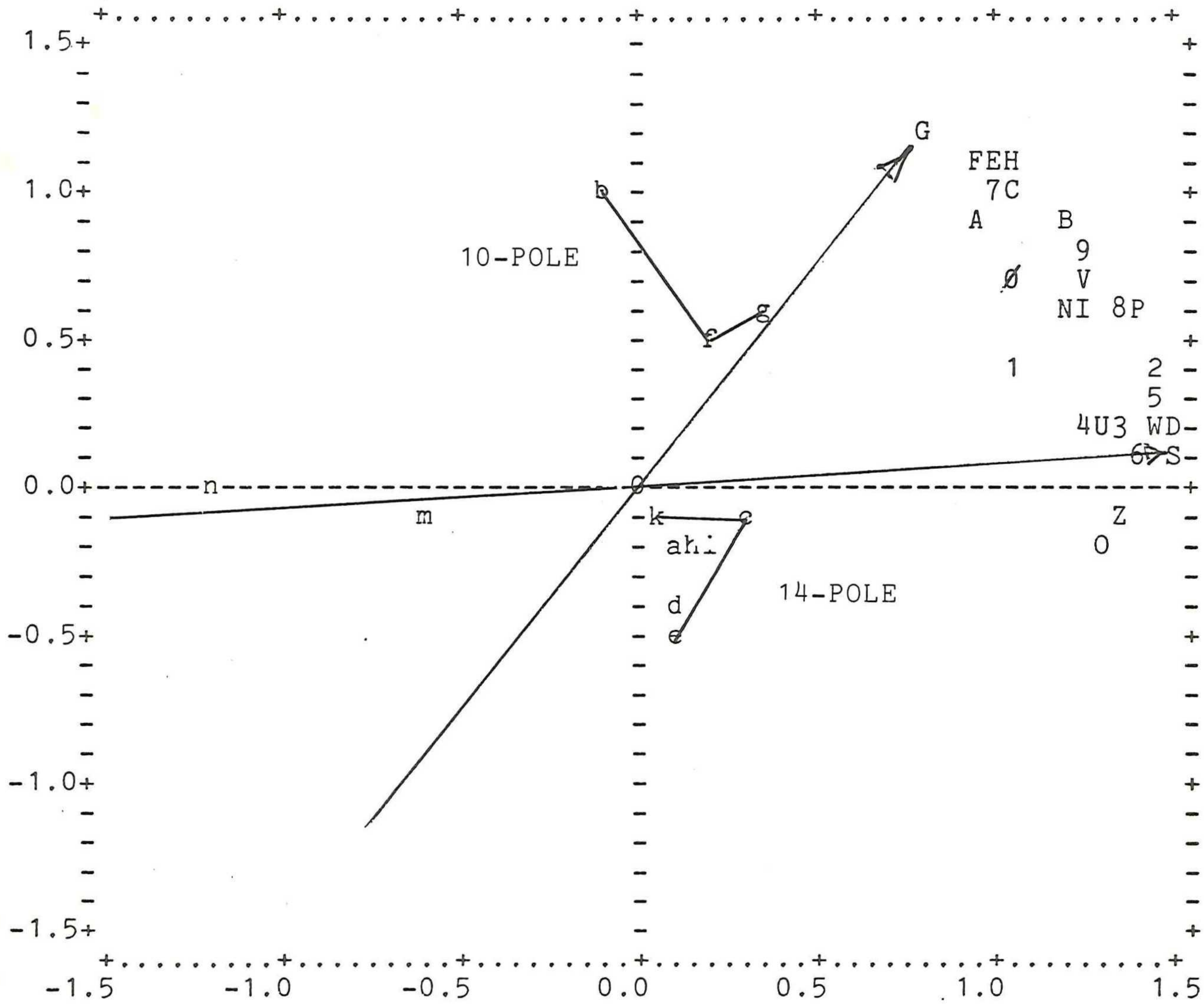
VECTOR 1 = SPKR 1 (AR), SENT 1.

VECTOR 4 = SPKR 1 (^R), SENT 4.

Fig. 1 - Appendix

MDPREF ANALYSIS OF DATA FILE: RTS3
REPLOT RATING DATA, 4S,36V,14P,3D; ROTATED PT2,3,2;PT14,1,2;PT14

PLOT OF SYSTEMS(PTS) AND SPKR*SENT VECTORS: DIM 1 VS DIM 3



DUPLICATE POINTS:

J=9, K=8, L=2, M=H, Q=5, R=4, T=6, X=5,

Y=D, j=i, l=h,

VECTOR G = SPKR 3 (DK), SENT 5

VECTOR S = SPKR 5 (RS), SENT 5.

Fig. 2 - Appendix